caBIG™ cancer Biomedical Informatics Grid™

caBIG™ Data Sharing Plan
Content Guideline

Draft
October 2007

PURPOSE

The caBIG™ (caBIG) Data Sharing Plan presents information about the data that the adopting institution will share via the caGrid, including how, with whom, and when the data will be shared. The plan establishes the nature of the data for purposes of determining the sensitivity of the data (related to legal and ethical restrictions on data exchange, intellectual property value, and contractual obligations) and thus the access controls necessary to secure the data. The data sharing plan also presents information about the institution's internal approval processes for sharing data outside of the institution. That information is used to deepen the understanding of the broader caBIG community about data sharing practices in general and to assist the individual adoption project and researcher in securing approval to share data.

The following list of topics to address is intended as a guideline or checklist to assist the researcher in preparing a Data Sharing Plan for use in a caBIG adoption project.

CONTENT OF THE DATA SHARING PLAN

I. Background of the Institution and the project that is adopting caBIG tools and using the caGrid to share data

- Describe the institution that is conducting the research (may be a unit or department of a larger institution):
  - Will any or all of the data be collected specifically for the current project?
  - If the data already exist, do any or all of the data to be shared fall within the Common Rule definition of human subjects research?
  - Is your institution a covered entity as defined by the Privacy Rule of the Health Insurance Portability and Accountability Act of 1996 (i.e., has your institution determined that it is subject to the requirements of the HIPAA Privacy Rule)?
- Describe the purpose and objectives of the project, the tools that will be adopted, and how the tools are expected to help achieve the purpose and objectives.
- Describe any sponsors (Federal, State, or private), collaborators, or others with regulatory, financial, intellectual property or other interests or rights to the data produced in the study or the data to be used in the study.

II. Information about the data involved in the project and how they will be shared

- Describe, with some specificity, the data that are intended to be shared, the source(s) of the data (e.g., mouse data, tissue bank specimens and annotations, imaged/annotated data, clinical trial or retrospective data from a study in progress, existing regular medical records, archived data).
- If the institution (or relevant component) is covered by the HIPAA Privacy Rule, state whether the health information to be shared includes direct identifiers, or only limited or de-identified data sets, as defined by the Rule (see chart below for guidance). If the dataset is de-identified, describe the process, whether manual or automated, used to de-identify it.

- Describe different expected levels of data access (e.g., public access, group or consortium access, or private access, which is generally limited to the originators of the data and/or individually authorized and authenticated individuals/institutions)..
- Describe the mechanism(s) for sharing the data such as whether the dataset will be a separate copy of data exposed to shared use outside of the institution's security firewall or the data will be accessed through the firewall.
- Describe any mechanisms in place within the institution (that is, mechanisms, if any, other than the authentication and authorization controls available either in the application itself or via the caGrid Common Security Module) that will be used for limiting access to authorized users. Note any differences where different types of data sets (as described above) may be involved.
- Describe what controls the institution expects to place on access to the data, if any and how authorized access will be administered (e.g., will there be a requirement for registration of individual users who must "sign" (electronically or physically) a data use agreement, either a standard "click-through" agreement presented by caGrid or a separately negotiated agreement between each prospective user and the research institution?)
- When do you expect to share the particular data relative to the close of the study (e.g., within N months after the close of the study to allow for data validation, immediately upon publication of the study which is expected to occur within N months of the close of the study, immediately upon submittal of a patent application which is expected to occur no later that N months after the close of the study), and by approximately what actual date would you expect the data to be shared?

III. Information about the institutional units involved in approval to share data outside of the institution, including an IRB

- Are there internal organizations other than an IRB involved in approving the adoption project and sharing data such as Technology Transfer Offices, Compliance, Privacy or Information Technology Security Offices, Legal Counsel, etc.? If so, describe the steps and approximate timeframes of the process for securing all necessary approvals, including what information the organization may require regarding the project, caGrid security or the caBIG program itself.
- If there is an IRB involved in approving data sharing, describe the steps and approximate timeframes involved in the process for securing approval to share the particular data involved in the adoption project, if any, including what information the organization may require regarding the project, caGrid security or the caBIG program itself.
- Does the IRB audit compliance with the options participants choose in the informed consent?

IV. Provide information about any additional anticipated challenges, limitations, or other constraints on data sharing.

caBIG™ cancer Biomedical Informatics Grid™

De-Identified Data Sets and Limited Data Sets

The following chart describes the information that must be *eliminated* from a database, registry, or any other data set for the data set to be considered "de-identified" or a "limited data set". Appropriately de-identified data sets are not regulated by the HIPAA Privacy Rule. Limited data sets may be used or disclosed for research, public health, and other limited purposes, but only by those who sign a "data use agreement" (signature may be written or electronic, depending on applicable state law and local institutional policy). Note that for each data element listed below, the information must be eliminated with respect to the patient *and* to any of the patient's relatives, employers, or household members.

Important: Even if the HIPAA Privacy Rule does not regulate the use of a dataset or permits its use or disclosure for research, federal regulations, institutional policies, and/or state laws and regulations governing human subjects research and health information privacy may still apply.

Data Element	De-Identified Data Set*	Limited Data Set
Names	Remove	Remove
Address, city and other geographic information smaller than state. *3-digit zip code may be included in a de-identified data set for an area where more than 20,000 people live; use "000" if fewer than 20,000 people live there.*	Remove	Remove postal address information other than city, town, state or zip code.
All elements of dates (except year); plus age and any date (including year) if age is over 89. *Examples: date of birth, date of death, date of admission, date of discharge, date of service.*	Remove	May be included.
Telephone, fax numbers; e-mail addresses, web URL addresses, IP addresses.	Remove	Remove
Social security number, medical record number, health plan beneficiary number, any account number, certificate or license number.	Remove	Remove
Vehicle identifiers and serial numbers, including license plate numbers.	Remove	Remove
Device identifiers and serial numbers.	Remove	Remove
Biometric identifiers (e.g., fingerprints; voice prints). *DNA is not considered a biometric identifier for purposes of HIPAA.*	Remove	Remove
Full-face photographs and any comparable images.	Remove	Remove
Any other unique identifying number, characteristic or code.	Remove†	May be included.

[*] Even if all of the information listed in this column is removed, if the researcher knows that any remaining information in the data set could be used to re-identify a patient (e.g., a diagnosis code where the disease is very rare), then the data set is not considered de-identified.

[†] If links must be maintained in the data set for potential later re-identification, they must be completely unrelated to any of the above elements. For example, a patient's initials or a scrambled social security number are not permitted in a de-identified data set. A subject code that reflects the order in which subjects were enrolled into a trial would be permitted.

Chart adapted from the original prepared by Rachel Nosowsky, Esq., University of Michigan., Rev. June 2003